

# Compatibilism

Michael McKenna

*First published Mon 26 Apr, 2004*

Compatibilism offers a solution to *the free will problem*. This philosophical problem concerns a disputed incompatibility between free will and determinism. *Compatibilism* is the thesis that free will is compatible with determinism. Because free will is taken to be a necessary condition of moral responsibility, compatibilism is sometimes expressed in terms of a compatibility between moral responsibility and determinism.

- [1. Terminology and One Formulation of the Free Will Problem](#)
  - [1.1 Free Will](#)
  - [1.2 Moral Responsibility](#)
  - [1.3 Determinism](#)
  - [1.4 Compatibilism's Competitors](#)
  - [1.5 The Free Will Problem](#)
- [2. Two Models of Control and Determinism's Apparent Threat to Free Will](#)
  - [2.1 Garden of Forking Paths Models and Alternative Possibilities Worries](#)
  - [2.2 Source Models and Source Worries](#)
  - [2.3 Compatibilists' Ameliorating Efforts](#)
- [3. Classical Compatibilism](#)
  - [3.1 Freedom According to Classical Compatibilism](#)
  - [3.2 The Lasting Influence of Classical Compatibilist Free Will](#)
  - [3.3 The Classical Compatibilist Conditional Analysis](#)
  - [3.4 The Lasting Influence of the Conditional Analysis](#)
- [4. Three Major Influences on Contemporary Compatibilism](#)
  - [4.1 The Consequence Argument](#)
  - [4.2 A Challenge to the Principle of Alternative Possibilities](#)
  - [4.3 Focus upon the Morally Reactive Attitudes](#)
- [5. Contemporary Compatibilism](#)
  - [5.1 Compatibilists' Responses to the Consequence Argument](#)
  - [5.2 Multiple Viewpoints Compatibilism](#)
  - [5.3 Hierarchical Compatibilism](#)
  - [5.4 The Reason View](#)
  - [5.5 Reasons-Responsive Compatibilism](#)
  - [5.6 Strawsonian Compatibilism](#)
- [Bibliography](#)
- [Other Internet Resources](#)
- [Related Entries](#)

---

## 1. Terminology and One formulation of the Free Will Problem.

## 1.1 Free Will

It would be misleading to specify a strict definition of free will since in the philosophical work devoted to this notion there is probably no single concept of it. For the most part, what philosophers working on this issue have been hunting for, maybe not exclusively, but centrally, is a feature of agency that is necessary for persons to be morally responsible for their conduct.<sup>[1]</sup> Different attempts to articulate the conditions for moral responsibility will yield different accounts of the sort of agency required to satisfy those conditions. What is needed, then, as a starting point, is a gentle, malleable notion that focuses upon special features of persons as agents. Hence, as a theory-neutral point of departure, free will can be defined as *the unique ability of persons to exercise control over their conduct in a manner necessary for moral responsibility*.<sup>[2]</sup> Clearly, this definition is too lean when taken as an endpoint; the hard philosophical work is about how best to develop this special kind of control. But however this notion of control is developed, its uniqueness consists, at least in part, in being possessed only by persons.

## 1.2 Moral Responsibility

A person who is a *morally responsible agent* is not merely a person who is able to *do* moral right or wrong. Beyond this, she is *accountable* for her morally significant conduct. Hence, she is, when fitting, an apt target of moral praise or blame, as well as reward or punishment. Free will is understood as a necessary condition of moral responsibility since it would seem unreasonable to say of a person that she deserves blame and punishment for her conduct if it turned out that she was not at any point in time in control of it. (Similar things can be said about praise and reward.) It is primarily, though not exclusively, because of the intimate connection between free will and moral responsibility that the free will problem is seen as an important one.<sup>[3]</sup>

## 1.3 Determinism

A standard characterization of determinism states that every event is causally necessitated by antecedent events.<sup>[4]</sup> Within this essay, we shall define determinism as the metaphysical thesis that *the facts of the past, in conjunction with the laws of nature, entail every truth about the future*. According to this characterization, if determinism is true, then, given the actual past, and holding fixed the laws of nature, only one future is possible at any moment in time. Notice that an implication of determinism as it applies to a person's conduct is that, if determinism is true, there are (causal) conditions for that person's actions located in the remote past, prior to her birth, that are sufficient for each of her actions.

## 1.4 Compatibilism's Competitors

The compatibilists' main adversaries are *incompatibilists*, who deny the compatibility of free will and determinism. Some incompatibilists remain agnostic as to whether persons have free will. But most take a further stand regarding the reality or unreality of free will. Some of these incompatibilists, *libertarians*, hold that at least some persons have free will and that, therefore, determinism is false. Other incompatibilists, *hard determinists*, have a less optimistic view, holding that determinism is true and that no persons have free will. A minority opinion is held by *hard incompatibilists*, who hold that there is no free will regardless of determinism's truth or falsity.

## 1.5 The Free Will Problem

If we are to understand compatibilism as a solution to the free will problem, it would be useful to have some sense of the problem itself. Unfortunately, just as there is no single notion of free will that unifies all of the work philosophers have devoted to it, there is no single specification of the free will problem. In fact, it might be more helpful to think in terms of a range of problems. Regardless, any formulation of the problem can be understood as arising from a troubling sort of entanglement of our concepts, an entanglement that seems to lead to contradictions, and thus that cries out for a sort of disentangling. In this regard, the free will problem is a classic philosophical problem; we are, it seems, committed in our thought

and talk to a set of concepts which, under scrutiny, appear to comprise a mutually inconsistent set. Formally, to settle the problem—to disentangle the set—we must either reject some concepts, or instead, we must demonstrate that the set is indeed consistent despite its appearance to the contrary. Just to illustrate, consider this set of propositions as an historically very well known means of formulating the free will problem. Call it the Classical Formulation:

1. Some person (qua agent), at some time, could have acted otherwise than she did.
2. Actions are events.
3. Every event has a cause.
4. If an event is caused, then it is causally determined.
5. If an event is an act that is causally determined, then the agent of the act could not have acted otherwise than in the way that she did.

This Classical Formulation involves principles governing six different concepts: a person as an agent, action, could have done otherwise, event, cause, and causal determination.<sup>[5]</sup> Note that this formulation involves a mutually inconsistent set of propositions, and yet each is (seemingly) rooted in our contemporary conception of the world. Proposition (1) is grounded in a conception of agency (free agency) and an understanding of ourselves as practical deliberators who are able to select amongst different possible courses of action. Proposition (2) is a partial definition of an action as something that takes place in time, something that can, for instance, have an identifiable duration. Proposition (3) is a presupposition of natural science. Proposition (4) is an operating assumption of natural science, or a considerable range of the natural sciences. And proposition (5) arises from a common sense understanding of what it means to claim that an event is causally determined—that, if it was, then given the antecedent causal conditions for the event, it was not possible for it not to have occurred.

By drawing upon the Classical Formulation, we can see how different stances might emerge. For instance, within the framework of this Classical Formulation, compatibilists would deny proposition (5). Incompatibilists, on the other hand, might move in a number of different directions. Consider the incompatibilist who remains agnostic about the free will problem. Her thesis is merely that free will and determinism are incompatible. Hence, given the Classical Formulation, she would be committed to the truth of proposition (5). Yet she is not prepared to say whether determinism is true or whether instead any person has free will. Her view is simply that there is no world in which it is the case that a person acts with freedom of the will and that world is determined. This sort of agnostic incompatibilist might frame her position by appeal to a disjunction, such as: Either (1) is false or (4) is false. (She might appeal to a different disjunction, such as: Either (1) is false or (3) is false.) Now consider the incompatibilist who commits to the hard determinist thesis that no person has free will and that determinism is true. Clearly, the hard determinist will reject proposition (1). Finally, consider the libertarian—the incompatibilist who embraces free will and denies that determinism is true. She has a number of options. She might deny (3), that every event is caused, thereby claiming that the universe is causally indeterministic. Or she might deny (4), that if an event is caused, then it is causally determined. On this account, she might well agree that actions are events, and that every event is caused (that is, she might accept propositions (2) and (3)), but she will claim that human agents are the cause of freely willed actions, and that human agents are not themselves caused (which would entail that they are not events).

The Classical Formulation of the free will problem has fallen out of fashion. It is meant to function here merely as an illustration of how different positions on the free will problem might emerge, and as an illustration of the ways that the differing positions might seek to disentangle the collection of concepts giving rise to the problem. To warn against settling exclusively on any single formulation of the free will problem, it might be instructive to show why this formulation is no longer helpful. Just to mention one problem with it, notice that the only proposition used to represent the freedom element of the notion of free will is proposition (1). However, as will become apparent later in this entry, there are notions of free will that do not appeal to a proposition involving the claim that an agent could have acted otherwise. All the same, such notions of free will do seem to be at odds with the thesis of causal determinism. Hence, there are debates between compatibilists and incompatibilists regarding a notion of free will that is entirely independent of *could have acted otherwise*.<sup>[6]</sup> In the absence of considerations having to do with an agent's

ability to act otherwise, the Classical Formulation does not provide the resources to show why free will might be thought to be incompatible with determinism.<sup>[7]</sup>

Rather than seek a formulation of the free will problem that allows a single, perspicuous demonstration of every possible position adopted with respect to it, it is more helpful to think in terms of different sorts of formulations. These different formulations will involve different considerations pertinent to the sort of freedom that is at issue when theorizing about the conditions for morally responsibility. In the following section, two formulations will be presented in the form of two arguments for incompatibilism. Regardless of the specific form they take, what is central to a proper understanding of them is that they emerge from an apparent problematic entangling of concepts that are a deep part of our conceptual repertoire. These concepts will include some subset of the following: freedom, control, person, agency, cause, causal necessity or determination, event, moral responsibility, as well as notions like the past, and a law of nature. The philosopher's task is to disentangle these various concepts in a useful and illuminating manner, seeking a set of true and consistent claims that captures as accurately as possible an understanding of ourselves as agents (allegedly) capable of acting of our own free will.

## 2. Two Models of Control and Determinism's Apparent Threat to Free Will

Determinism poses at least two different sorts of threats to free will. In each case, we can begin with the theory-neutral definition of free will set out in section one: *the unique ability of persons to exercise control over their conduct in a manner necessary for moral responsibility*. This characterization of free will in terms of control can be developed in two ways. One concerns an agent's freedom over alternatives. Another concerns the source of an agent's actions. Incompatibilists have rightly exploited each. Each builds upon different models of control, and each has instigated different incompatibilist formulations of the free will problem.

### 2.1 Garden of Forking Paths Models and Alternative Possibilities Worries

A natural way to think of an agent's control over her conduct at a moment in time is in terms of her ability to select among, or choose between, alternative courses of action.<sup>[8]</sup> This picture of control stems from common features of our perspectives as practical deliberators settling on courses of action. If one is choosing between voting for Gore as opposed to Bush, it is plausible to assume that her freedom with regard to her voting consists, at least partially, in her ability to choose between these two alternatives. On this account, acting with free will requires *alternative possibilities*. A natural way to model this account of free will is in terms of an agent's future as a garden of forking paths branching off from a single past. A locus of freely willed action arises when the present offers, from an agent's (singular) past, more than one path into the future. Borrowing from the Argentine fabulist Borges, let us call this the *Garden of Forking Paths* model of control.<sup>[9]</sup> Let us say, as the Garden of Forking Paths suggests, that when a person acts of her own free will, *she could have acted otherwise*. Her ability to have acted otherwise is underwritten by her ability to have selected amongst, or chosen between, alternative courses of action.

Unpacking free will by appeal to the Garden of Forking Paths model immediately suggests that determinism might be a threat to it. For determinism, understood in the strict sense characterized above, tells us that, at any time, given the facts of the past and the laws of nature, *only* one future is possible. But the Garden of Forking Paths model suggests that a freely willing agent could have acted other than she did and, hence, that *more* than one future is possible.

Here is an incompatibilist argument that codifies the considerations set out above:

- A. Any agent,  $x$ , performs any act  $a$  of  $x$ 's own free will iff  $x$  has control over  $a$ .
- B.  $x$  has control over  $a$  only if  $x$  has the ability to select among alternative courses of action to act  $a$ .
- C. If  $x$  has the ability to select among alternative courses of action to act  $a$ , then there are alternative courses of action to act  $a$  open to  $x$  (i.e.,  $x$  could have done otherwise than  $a$ ).

- D. If determinism is true, then only one future is possible given the actual past, and holding fixed the laws of nature.
- E. If only one future is possible given the actual past, and holding fixed the laws of nature, then there are no alternative courses of action to any act open to any agent (i.e., no agent could have done otherwise than she actually does).
- F. Therefore, if determinism is true, it is not the case that any agent,  $x$ , performs any act,  $a$ , of her own free will.<sup>[10]</sup>

For ease of reference and discussion throughout this entry, let us simplify the above argument as follow:

- 1. If a person acts of her own free will, then she could have done otherwise (A-C).
- 2. If determinism is true, no one can do otherwise than one actually does (D-E).
- 3. Therefore, if determinism is true, no one acts of her own free will (F).

Call this simplified argument the *Classical Incompatibilist Argument*. According to the argument, if determinism is true, no one has access to alternatives in the way required by the Garden of Forking Paths model.<sup>[11]</sup>

## 2.2 Source Models and Source Worries

Here is a different way to develop the notion of control. An agent's control consists in her playing a crucial role in the production of her actions.<sup>[12]</sup> Think in terms of the transparent difference between those events that are *products of one's agency* and those that are merely bodily happenings. For instance, consider the choice to pick up a cup of coffee as opposed to the event of one's heart beating or one's blood circulating. In the latter cases, one recognizes events happening to one; in the former, one is the source and producer of that happening. Control is understood as one's being the source whence her actions emanate. On this model, a *Source* model of control, one's actions issue from one's self (in a suitable manner).

Fixing just upon the Source model, how might determinism pose a threat to free will? If determinism is true, then for any person, there are facts of the past prior to her birth that, when combined with the laws of nature, provide causally sufficient conditions for the production of her actions. But if this is so, then, while it might be true that an agent is indeed a source of her action, that source *originates* outside of her. Hence, she, as an agent, is not the *ultimate source* of her actions. What is meant here by an ultimate source, and not just a source? When an agent is an ultimate source of her action, some condition necessary for her action originates with the agent herself. It cannot be located in places and times prior to the agent's freely willing her action. If an agent is not the ultimate source of her actions, then her actions do not originate in her, and if her actions are the outcomes of conditions guaranteeing them, how can she be said to control them? The conditions sufficient for their occurrence were already in place long before she even existed!

Here is an incompatibilist argument that codifies the considerations set out above:

- A. Any agent,  $x$ , performs an any act,  $a$ , of her own free will iff  $x$  has control over  $a$ .
- B.  $x$  has control over  $a$  only if  $x$  is the ultimate source of  $a$ .
- C. If  $x$  is the ultimate source of  $a$ , then some condition,  $b$ , necessary for  $a$ , originates with  $x$ .
- D. If any condition,  $b$ , originates with  $x$ , then there are no conditions sufficient for  $b$  independent of  $x$ .
- E. If determinism is true, then the facts of the past, in conjunction with the laws of nature, entail every truth about the future.
- F. If the facts of the past, in conjunction with the laws of nature, entail every truth about the future, then for any condition,  $b$ , necessary for any action,  $a$ , performed by any agent,  $x$ , there are conditions independent of  $x$  (in  $x$ 's remote past, before  $x$ 's birth) that are sufficient for  $b$ .
- G. If, for any condition,  $b$ , necessary for any action,  $a$ , performed by any agent,  $x$ , there are conditions independent of  $x$  that are sufficient for  $b$ , then no agent,  $x$ , is the ultimate source of any action,  $a$ . (This follows from C and D.)

- H. If determinism is true, then no agent, *x*, is the ultimate source of any action, *a*. (This follows from E, F, and G.)
- I. Therefore, if determinism is true, then no agent, *x*, performs any action, *a*, of her own free will. (This follows from A, B, and H.)<sup>[13]</sup>

For ease of reference and discussion throughout this entry, let us simplify the above argument as follow:

- 1. A person acts of her own free will only if she is its ultimate source (A-B).
- 2. If determinism is true, no one is the ultimate source of her actions (C-H).
- 3. Therefore, if determinism is true, no one acts of her own free will (I).

Call this simplified argument the *Source Incompatibilist Argument*. It is important to see that the demand for alternative possibilities as illustrated on the Garden of Forking Paths Model is not (at least not obviously) relevant to this incompatibilist argument. Suppose, as the Garden of Forking Path Model suggests, that a putatively freely willing agent had access to the relevant sort of alternative possibilities. According to the Source Incompatibilist Argument, a *further* condition is that she must have been the ultimate source of her freely willed actions. Furthermore, even if, for some reason, agency of the sort indicated by the Garden of Forking Paths model were not necessary for free will, the Source Incompatibilist Argument would carry independent force. Hence, grant for the sake of argument that it is possible for an agent to act of her own free will without the relevant sort of alternative possibilities. According to the Source Incompatibilist Argument, for an agent to take the particular path that she takes and in doing so act of her own free will, she has to be the *ultimate* source of that path. If determinism is true, then, while it might appear to an agent that she plays a role in the production of her action, her contribution to the subsequent action is not significant in the way required for her to act of her own free will. Why is it not significant? What explains why she acts need make no reference to her.<sup>[14]</sup>

### 2.3 Compatibilists' Ameliorating Efforts

In assessing compatibilist theories and arguments, it is useful consider what sort of model of control they rely upon—Garden of Forking Paths or Source—and how they stack up against both the Classical Incompatibilist Argument and the Source Incompatibilist Argument. As for the Classical Incompatibilist Argument, some compatibilists have responded to this argument by denying the truth of the second premise: If determinism is true, no one can do otherwise than one actually does. By doing so, these compatibilists embrace a Garden of Forking Paths model of control. They maintain that determinism is not a threat to it. (For example, see sections 3.3, 5.1, and 5.4.) Others have instead resisted the first premise: If a person acts of her own free will, then she could have done otherwise. These compatibilists proceed by rejecting the Garden of Forking Paths model altogether. (See sections 4.2, 5.2, 5.3, 5.5, and 5.6.) They instead attempt to make do with a Source model of control. What, then, of the Source Incompatibilist Argument? No compatibilist, it seems, can deny the truth of the second premise of the Source Incompatibilist Argument: If determinism is true, no one is the ultimate source of her actions. Given the definition of ultimacy, the second premise amounts to an analytic truth. Thus, all compatibilists must respond to the argument by arguing against the truth of the first premise: A person acts of her own free will only if she is its ultimate source.

Let us turn now from incompatibilism's plausible concerns to compatibilism itself.

## 3. Classical Compatibilism

A useful manner of thinking about compatibilism's place in contemporary philosophy is in terms of at least three stages. The first stage involves the classical form defended in the modern era by the empiricists Hobbes and Hume, and reinvigorated in the early part of the twentieth century. The second stage involves three distinct contributions in the 1960s, contributions that challenged many of the dialectical presuppositions driving classical compatibilism. The third stage involves various contemporary forms of compatibilism, forms that diverge from the classical variety and that emerged out of, or resonate with, at

least one of the three contributions found in the second transitional stage. This section will be devoted to the first stage, to that of classical compatibilism.

Classical compatibilism is associated with several distinct theses. Only two will be considered here. One involves a strikingly austere account of freedom. A second involves an attempt to explain how an agent could be free to do otherwise even if she was determined to do what she did.

### 3.1 Freedom According to Classical Compatibilism

According one strand within classical compatibilism, freedom of the sort pertinent to moral evaluation is nothing more than an agent's ability to do what she wishes in the absence of impediments that would otherwise stand in her way. For instance, Hobbes writes that a person's freedom consists in his finding, "no stop, in doing what he has the will, desire, or inclination to do" (*Leviathan*, p.108). Hobbes' brief remarks represent an exemplary expression of the classical compatibilist account of freedom. It involves two components, a positive and a negative one. The positive component (doing what one wills, desires, or inclines to do) consists in nothing more than what is involved in the power of agency. The negative component (finding "no stop") consists in acting unencumbered or unimpeded. Typically, the classical compatibilists' benchmark of impeded or encumbered action is compelled action. Compelled action arises when one is forced by some foreign or external source to act contrary to her will.

Classical compatibilism is often associated with the thesis that the word *freedom* in the expression *freedom of will* modifies a condition of action and not will. For this reason, some writers advised burying the expression altogether and instead speaking only in terms of *freedom of action* (e.g., Schlick, 1939). For ease of expression, and to avoid cumbersome worries about different authors' formulations, let us characterize the moral freedom pertinent to classical compatibilism as freedom of will, keeping in mind that this notion is meant to be a deflationary one attributing nothing special to the will itself.

Free will is the unencumbered ability of an agent to do what she wants.

It is plausible to assume that free will, so understood, is compatible with determinism since the truth of determinism does not entail that no agents ever do what they wish to do unencumbered.

How convincing is the classical compatibilist account of free will? As it stands, it cries out for refinement. To cite just one shortcoming, various mental illnesses can cause a person to act as she wants, to do so unencumbered, and yet intuitively, it would seem that she does not act of her own free will. For example, imagine a person suffering from a form of psychosis that causes full-fledged hallucinations. While hallucinating, she might "act as she wants unencumbered," but she could hardly be said to be acting of her own free will.

### 3.2 The Lasting Influence of Classical Compatibilist Free Will

The classical compatibilist account of free will (the unencumbered ability to do as one wants) permits all kinds of cases that intuitively seem to involve agents lacking freedom of will. Sometimes the very wants giving rise to actions are the sources of an agent's lack of freedom, as in cases of addiction or neuroses. But perhaps the wants leading to freely willed actions can be more carefully captured in ways that fit the spirit of the classical compatibilists' strategy. If somehow those wants could be more narrowly specified so as to rule out the deviant freedom-undermining ones, then, like the classical compatibilist, some brand of compatibilism could show that simple, uncontroversial features of agency, or maybe of an agent's deliberative capacities, are adequate to capture the kind of freedom required for freedom of will. In subsequent sections, we will see that several contemporary compatibilist efforts adopt this approach, an approach instigated by way of the classical compatibilist account of free will.

The classical compatibilist account of free will, even if incomplete, can be contrasted with the Source Incompatibilist Argument discussed in section 2.2. The dispute is over the truth of the first premise of that

argument: A person acts of her own free will only if she is its ultimate source. No doubt, for one to be an ultimate source of her action, no explanation for her action can trace back to factors prior to her. This the compatibilist cannot have since it requires the falsity of determinism. But according to the classical compatibilist account of free will, so long as one's action arises from one's unencumbered desires, she is a genuine source of her action. Surely she is not an ultimate, only a mediated source. But she is a source all the same, and this sort of source of action, the classical compatibilist will argue, is sufficient to satisfy the kind of freedom required for free will and moral responsibility. This general classical compatibilist strategy—developing an appropriately nuanced account of the source of agency—offers a lasting contribution to the free will debate. Contemporary compatibilist variations must adopt some similar posture towards the Source Incompatibilist Argument.

### 3.3 The Classical Compatibilist Conditional Analysis

Consider the following incompatibilist objection to the classical compatibilist account of free will:

If determinism is true, and if at any given time, an unencumbered agent is completely determined to have the wants that she does have, and if those wants causally determine her actions, then, even though she does do what she wants to do, *she can not ever do otherwise*. She satisfies the classical compatibilist conditions for free will. But free will requires the ability to do otherwise, and determinism is incompatible with this. Hence, the classical compatibilist account of free will is inadequate. Determinism is incompatible with free will and moral responsibility because determinism is incompatible with the ability to do otherwise.

Is the incompatibilist correct? Notice that the incompatibilist's objection does not deny either of the classical compatibilist conditions of free will. Her concern is, even if they are necessary, they are not jointly sufficient; a freely willing agent must also be able to do otherwise.

Some classical compatibilists, such as Hobbes, did not take up this incompatibilist challenge, opting for a sort of *one-way* freedom that required only that what a person did do, she did unencumbered as an upshot of her own agency. Thus, the one-way classical compatibilists relied exclusively on a Source Model of control. But other classical compatibilists took the challenge seriously and argued for a sort of *two-way* freedom.<sup>[15]</sup> Clearly, the two-way compatibilists were prepared to defend a Garden of Forking Paths model of control.

The two-way classical compatibilists responded by arguing that determinism *is* compatible with the ability to do otherwise. To show this, they attempted to analyze an agent's ability to do otherwise in conditional terms (e.g., Hume, *Enquiry Concerning Human Understanding*, p.73; Ayer, 1954; or Hobart, 1934). Since determinism is a thesis about what must happen in the future *given the actual past*, determinism is consistent with the future being different *given a different past*. So the classical compatibilists analyzed any assertion that an agent could have done otherwise as a conditional assertion reporting what an agent would have done under certain counterfactual conditions. These conditions involved variations on what a freely willing agent wanted (chose, willed, or decided) to do at the time of her freely willed action. Suppose that an agent freely willed *X*. According to the classical compatibilist *conditional analysis*, to say that, at the time of acting, she could have done *Y* and not *X* is just to say that, *had* she wanted (chosen, willed, or decided) to do *Y* and not *X* at that time, *then* she would have done *Y* and not done *X*. Her ability to have done otherwise at the time at which she acted consisted in some such counterfactual truth.

Is this analysis plausible? Given that a determined agent is determined at the time of action to have the wants that she does have, how is it helpful to state what she would have done had she had different wants than the wants that she did have? Assuming the truth of determinism, at the time at which she acted, according to the incompatibilist, she could have had no other wants than the wants that her causal history determined her to have. How is this counterfactual ability more than a hollow freedom? The classical compatibilist held that the conditional analysis brings into relief a rich picture of freedom. In assessing an agent's action, the analysis accurately distinguishes between those actions she would have performed if she wanted, from those actions she could not have performed even if she wanted. This, the classical

compatibilist held, effectively distinguishes between those alternative courses of action that were within the scope of the agent's abilities at the time of action, from those courses of action that were not. This just is the distinction between what an agent was free to do from what she was not free to do. This is not at all a hollow freedom; it demarcates what persons have within their control from what falls outside that purview.

Despite the classical compatibilists' ingenuity, their analysis of *could have done otherwise* failed decisively. The classical compatibilists wanted to show their incompatibilist interlocutors that when one asserted that a freely willing agent had alternatives available to her—that is, when it was asserted that she could have done otherwise—that assertion could be analyzed as a conditional statement, a statement that is perspicuously compatible with determinism. But as it turned out, the analysis was refuted when it was shown that the conditional statements sometimes yielded the improper result that a person was able to do otherwise even though it was clear that at the time the person acted, she had no such alternative and therefore was not able to do otherwise in the pertinent sense (Chisholm, 1964, in Watson, ed., 1982, pp.26-7; or van Inwagen, 1983, pp.114-9). Here is such an example:

Suppose that Danielle is psychologically incapable of wanting to touch a blond haired dog. Imagine that, on her sixteenth birthday, unaware of her condition, her father brings to her two puppies to choose between, one being a blond haired Lab, the other a black haired Lab. He tells Danielle just to pick up whichever of the two she pleases and that he will return the other puppy to the pet store. Danielle happily, and unencumbered, does what she wants and picks up the black Lab.

When Danielle picked up the black Lab, was she able to pick up the blond Lab? It seems not. Picking up the blond Lab was an alternative that was not available to her. In this respect, *she could not have done otherwise*. Given her psychological condition, she cannot even form a want to touch a blond Lab, hence she could not pick one up. But notice that, *if she wanted to pick up the blond Lab, then she would have done so*. Of course, if she wanted to pick up the blond Lab, then she would not suffer from the very psychological disorder that causes her to be unable to pick up blond haired doggies. The classical compatibilist analysis of 'could have done otherwise' fails. According to the analysis, when Danielle picked up the black Lab, she *was* able to pick up the blonde Lab, even though, due to her psychological condition, she *was not* able to do so in the relevant respect. Hence, the analysis yields the wrong result.

The classical compatibilist attempt to answer the incompatibilist objection failed. Even if an unencumbered agent does what she wants, if she is determined, at least as the incompatibilist maintains, she could not have done otherwise. Since, as the objection goes, freedom of will requires freedom involving alternative possibilities, classical compatibilist freedom falls.

### 3.4 The Lasting Influence of the Conditional Analysis

It should be pointed out that the classical compatibilists' *failure* to prove that 'could have done otherwise' statements *are* compatible with determinism does not amount to a proof that 'could have done otherwise' statements are *incompatible* with determinism. So the incompatibilists' compelling counterexamples to the analysis (such as the one involving Danielle and the blond haired puppy) do not alone prove that determinism is incompatible with the freedom to do otherwise.

Despite this qualification, given the classical compatibilists' failure, they had no reply to the Classical Incompatibilist Argument. What the classical compatibilists attempted to do by way of their conditional analysis was deny the truth of the second premise: If determinism is true, no one can do otherwise. But, given their failure, it was incumbent upon them to respond to the argument in some manner. It is only dialectically fair to acknowledge that determinism does pose a *prima facie* threat to free will when free will is understood in terms of the Garden of Forking Paths model. The Classical Incompatibilist Argument is merely a codification of this natural thought. In light of the failure of the classical compatibilists' conditional analysis, the burden of proof rests squarely on the compatibilists. How can the freedom to do otherwise be reconciled with determinism? It is the mark of contemporary compatibilists that they speak to this issue.

#### 4. Three Major Influences on Contemporary Compatibilism

In the 1960's, three major contributions to the free will debate radically altered it. One was an incompatibilist argument that put crisply the intuition that a determined agent does not have control over alternatives. This argument, first developed by Carl Ginet, came to be known as the *Consequence Argument* (Ginet, 1966). Another contribution was Harry Frankfurt's argument against the *Principle of Alternative Possibilities* (PAP), a principle stating that an agent is morally responsible for what she does only if she can do otherwise (Frankfurt, 1969). Frankfurt's argument turned upon an example in which, it was argued, an agent could not do otherwise, but, intuitively, the agent was morally responsible. Finally, P.F. Strawson defended compatibilism by inviting both compatibilists and incompatibilists to attend more carefully to the central role of the *morally reactive attitudes* in understanding the concept of moral responsibility (Strawson, 1962). According to Strawson, the threat determinism allegedly poses to free will and moral responsibility is defused once the place of the morally reactive attitudes is properly appreciated. Each of these contributions changed dramatically the way that the free will problem is addressed in contemporary discussions. No account of free will, compatibilist or incompatibilist, is advanced today without taking into account at least one of these three pieces.

##### 4.1 The Consequence Argument

This argument invokes a compelling pattern of inference regarding claims about what is *power necessary* for a person. Power necessity, as applied to true propositions (or facts), concerns what is *not* within a person's power. Or, put differently, it concerns facts that a person does not have power over. *To say that a person does not have power over a fact is to say that she cannot act in such a way that the fact would not obtain.* To illustrate, no person has power over the truths of mathematics. That is, no person can act in such a way that the truths of mathematics would be false.<sup>[16]</sup> Hence, the truths of mathematics are, for any person, power necessary. Intuitively, the pattern of inference applied to these claims is simply that if a person has no power over a certain fact, and if she also has no power over the further fact that the original has some other fact as a consequence, then she also has no power over the consequent fact. Powerlessness, it seems, transfers from one fact to consequences of it. For example, if poker-playing Diamond Jim, who is holding only two pairs, has no power over the fact that Calamity Sam draws a straight flush, and if a straight flush beats two pairs (and assuming Jim has no power to alter this fact), then it follows that Jim has no power over the fact that Sam's straight flush beats Jim's two pairs. This general pattern of inference is applied to the thesis of determinism to yield a powerful argument for incompatibilism. The argument requires the assumption that determinism is true, and that the facts of the past and the laws of nature are fixed. Given these assumptions, here is a rough, non-technical sketch of the argument:<sup>[17]</sup>

1. No one has power over the facts of the past and the laws of nature.
2. No one has power over the fact that the facts of the past and the laws of nature entail that only one future is possible (i.e., determinism is true).
3. Therefore, no one has power over the facts of the future.

According to the Consequence Argument, *if* determinism is true, it appears that no person has any power to alter how her own future will unfold.

The Consequence Argument shook compatibilism, and rightly so. The classical compatibilists' failure to analyze statements of an agent's abilities in terms of counterfactual conditionals (see section 3.3) left the compatibilists with no perspicuous retort to the crucial second premise of the Classical Incompatibilist Argument: If determinism is true, no one can do otherwise (see section 2.1). The Consequence Argument, on the other hand, offered the incompatibilists powerful support of this second premise. If, according to the consequence argument, determinism implies that the future will unfold in only one way, and if no one has any power to alter its unfolding in that way, then it seems that, in a very clearly presented manner, no one can do other than she does. It is fair to say that the Consequence Argument earned the incompatibilists the dialectical advantage. The burden of proof was placed upon the compatibilists, at least to show what was wrong with the Consequence Argument, and better yet, to provide some positive account of the ability to

do otherwise. Seemingly, the compatibilists' only way around this burden was to defend compatibilism without relying upon the freedom to do otherwise.

#### 4.2 A Challenge to the Principle of Alternative Possibilities

As suggested above (section 4.1), one compatibilist strategy is to sidestep the debate over the truth of the second premise of the Classical Incompatibilist Argument as set out in section 2.1: If determinism is true, no one can do otherwise. An alternative strategy is to attack the first premise of the Classical Incompatibilist Argument: If a person acts of her own free will, then she could have done otherwise. In this way, compatibilists would turn away from the Garden of Forking Paths model of control and seek some other model as the intuitive basis for the kind of control pertinent to morally responsible agency. In his seminal 1969 paper, "Moral Responsibility and Alternate Possibilities," Harry Frankfurt developed an argument that gave compatibilists the resources to argue in just this way. Frankfurt's argument was directed against the Principle of Alternative Possibilities (PAP):

*PAP*: A person is morally responsible for what she does do only if she can do otherwise.

Central to Frankfurt's attack on PAP is a type of example in which an agent is morally responsible, but could not, at the time of the pertinent action, do otherwise. Here is a close approximation to the example Frankfurt presented in his original paper:

Jones has resolved to shoot Smith. Black has learned of Jones' plan and wants Jones to shoot Smith. But Black would prefer that Jones shoot Smith on his own. However, concerned that Jones might waiver in his resolve to shoot Smith, Black secretly arranges things so that, if Jones should show any sign at all that he will not shoot Smith (something Black has the resources to detect), Black will be able to manipulate Jones in such a way that Jones will shoot Smith. As things transpire, Jones follows through with his plans and shoots Smith for his own reasons. No one else in any way threatened or coerced Jones, offered Jones a bribe, or even suggested that he shoot Smith. Jones shot Smith under his own steam. Black never intervened.

In this example, Jones shot Smith on his own, and did so unencumbered — did so freely. But, given Black's presence in the scenario, Jones could not have done otherwise than shoot Smith. Hence, we have a counterexample to PAP.

If Frankfurt's argument against PAP is correct, the free will debate has been systematically miscast through much of the history of philosophy. *If* determinism threatens free will and moral responsibility, it is *not* because it is incompatible with the ability to do otherwise. Even if determinism *is* incompatible with a sort of freedom involving the ability to do otherwise, *it is not the kind of freedom required for moral responsibility*.

An enormous literature has emerged around the success of Frankfurt's argument and, in particular, around the example Frankfurt offered as contrary to PAP.<sup>[18]</sup> The debate is very much alive, and no clear victor has emerged (in the way that the incompatibilists can rightly claim to have laid to rest the compatibilists' conditional analysis strategy (see section 3.3)). Regardless, what is most relevant to this essay is that Frankfurt's argument instigated many compatibilists to begin thinking about accounts of freedom or control that unabashedly turn away from a Garden of Forking Paths model.

#### 4.3 Focus upon the Morally Reactive Attitudes

In "Freedom and Resentment" (1962), P.F. Strawson broke ranks with the classical compatibilists. Strawson developed three distinct arguments for compatibilism, arguments quite different from those the classical compatibilists endorsed. But more valuable than his arguments was his general theory of what moral responsibility is, and hence, what is at stake in arguing about it. Strawson held that both the incompatibilists and the compatibilists had misconstrued the nature of moral responsibility. Having lost

sight of it, they had failed to appreciate fully what an indictment of it would come to. All parties, Strawson suggested, advanced arguments in support of or against a distorted simulacra of the real deal.

#### **4.3.1 Strawson's Theory of Moral Responsibility**

To understand moral responsibility properly, Strawson invited his reader to consider the reactive attitudes one has towards another when she recognizes in another's conduct an attitude of ill will. The reactions that flow naturally from witnessing ill will are themselves attitudes that are directed at the perpetrator's intentions or attitudes. When a perpetrator wrongs a person, she, the wronged party, typically has a personal reactive attitude of resentment. When the perpetrator wrongs another, some third party, the natural reactive attitude is moral indignation, or disapprobation, which amounts to a “vicarious analogue” of resentment felt on behalf of the wronged party. When one is oneself the wronging party, reflecting upon or coming to realize the wrong done to another, the natural reactive attitude is guilt.

Strawson wanted contestants to the free will debate to see more clearly than they had that excusing a person — electing not to hold her morally responsible — involves more than some objective judgment that she did not do such and such, or did not intend so and so, and therefore does not merit some treatment or other. It involves a suspension or withdrawal of certain morally reactive attitudes, attitudes involving emotional responses. On Strawson's view, what it is to hold a person morally responsible for wrong conduct *is* nothing more than the propensity towards, or the sustaining of, a morally reactive attitude of disapprobation. Crucially, the disapprobation is in response to the perceived attitude of ill will or culpable motive in the conduct of the person being held responsible. Hence, Strawson explains, posing the question of whether the entire framework of moral responsibility should be given up as irrational (if it were discovered that determinism is true), is tantamount to posing the question of whether persons in the interpersonal community — that is, in real life — should forswear having reactive attitudes towards persons who wrong others, and who sometimes do so intentionally. Strawson invites us to see that the morally reactive attitudes that are the constitutive basis of our moral responsibility practices, as well as the interpersonal relations and expectations that give structure to these attitudes, are deeply interwoven into human life. These attitudes, relations and expectations are so much an expression of natural, basic features of our social lives — of its emotional texture — that it is virtually inconceivable to imagine how they *could* be given up.

Given this brief sketch of Strawson's theory of moral responsibility, let us consider in its light just two of Strawson's three arguments for compatibilism.

#### **4.3.2 Strawson's Psychological Impossibility Argument**

Strawson argued that it would be psychologically impossible to stop holding persons morally responsible for their conduct since it would be psychologically impossible to stop having certain kinds of emotional responses to others, responses that are simply part of our very nature. Hence, arguing about whether or not determinism threatens moral responsibility is idle. There is no way to take seriously the threat that it does, since, were it taken to discredit moral responsibility, given the nature of the human condition, no one could simply opt out of all moral responsibility practices anyway.

#### **4.3.3 Strawson's Practical Rationality Argument**

Strawson also argued that, *per impossible*, were we to be able to give up the morally reactive attitudes, and hence, give up all moral responsibility practices, and if determinism were to show that moral responsibility is grounded on falsehoods, even then, the only rational basis for deciding whether or not to forsake the reactive attitudes altogether would be in terms of the gains and losses to human life. And, Strawson suggests, the richness brought to human life by seeing persons as members of an interconnected community, and hence as morally responsible for their conduct, would far outweigh any other benefits that could be gained by giving up these practices.

#### 4.3.4 The Lasting Influence of Strawsonian Compatibilism

Strawson's arguments for compatibilism have not evaded scrutiny.<sup>[19]</sup> But regardless of their soundness, the deep insight he offered as to the nature of moral responsibility has pervaded the contemporary debate, and it should have. Strawson gave philosophers a clear picture of the phenomena under discussion, and of its importance in human life. Strawson made obvious to all that moral responsibility, as a genuine human practice, involves attitudes, emotions, and most crucially, a deference to the perspective of those holding agents morally responsible, that is, to those prone to the reactive attitudes. That perspective and those attitudes can be used to shine a light on the conditions for morally responsible agency and, especially, on the kind of freedom compatibilists and incompatibilists are concerned to discuss.

### 5. Contemporary Compatibilism

As set out in the previous section (Section 4), three major contributions in the 1960's profoundly altered the face of compatibilism: the incompatibilists' Consequence Argument (Section 4.1), Frankfurt's attack on the Principle of Alternative Possibilities (PAP) (Section 4.2), and Strawson's focus upon the morally reactive attitudes (Section 4.3). Each instigated major developments in contemporary debates about free will. Every account of compatibilism in the contemporary literature is shaped in some way by at least one of these influences. This section will focus upon six of the most significant contemporary compatibilist positions. Those wishing to learn about cutting edge work can read the supplement on

#### [Compatibilism: The State of the Art](#)

Before considering any particular contemporary compatibilist position, it is worth calling attention to one important distinction. Some contemporary compatibilist strategies attempt to capture freedom in terms of alternative possibilities; others do not. Frankfurt (1971) drew a distinction between *acting with a will that is free* and *acting of one's own free will*, the former requiring alternative possibilities, the latter not requiring them. But a more useful bit of terminology was introduced by John Martin Fischer (1982, 1994). As Fischer's has it, an agent with *regulative control* can, so to speak, regulate between different alternatives. An agent with *guidance control* guides or brings about her conduct even if she has no other alternatives to the course she takes.

As Fischer points out, an agent could possess both guidance and regulative control, but the two can come apart. On a view like Fischer's or Frankfurt's, it is only guidance control that is necessary for moral responsibility. Frankfurt's attack on PAP (see Section 4.2) prompted many contemporary compatibilists to develop accounts of compatibilist freedom that make no appeal to regulative control as modeled on a Garden of Forking Paths. Such accounts of guidance control fix solely on Source models of control, showing that an agent plays a special sort of role in the actual bringing about of her freely willed actions. Other compatibilists retained the classical compatibilist commitment to show that determined agents are able to act with regulative control.

### 5.1 Compatibilists' Responses to the Consequence Argument

The Consequence Argument (section 4.1) makes a strong case for the incompatibility of determinism and freedom involving alternative possibilities. It states that:

1. No one has power over the facts of the past and the laws of nature.
2. No one has power over the fact that the facts of the past and the laws of nature entail that only one future is possible (i.e., determinism is true).
3. Therefore, no one has power over the facts of the future.

Compatibilists defending a Garden of Forking Paths model of regulative control must show what is wrong with this powerful argument. Let us consider three different compatibilist attempts to unseat it.<sup>[20]</sup>

### 5.1.1 Challenging Power Necessity and the Past

Some compatibilists have argued against the first premise of the Consequence Argument by attempting to show that a person *can* act in such a way that the past would be different. Consider the difference between a person in the present who has the ability to act in such a way that *she alters the past*, as opposed to a person who has the ability to act in such a way such that, *if she did so act, the past would have been different*. Notice that the former ability is outlandish; it would require magical powers. But the latter ability is, at least by comparison, uncontroversial. It merely indicates that a person who acted a certain way at a certain time possessed abilities to act in various sorts of ways. Had she exercised one of those abilities, and thereby acted differently, then the past leading up to her action would have been different. To illustrate how comparatively mild such a claim about an agent's ability and the past might be, think about a logically similar sort of claim that is simply about what would be required for an agent to act differently. For example, consider the claim, *If I were dancing on the French Riviera right now, I'd be a lot richer than I am*. Certainly this claim does not mean (at least not given my dancing skills) that if I go to the French Riviera to dance, I will *thereby* be made richer. It only means that were I to have gone there to tango, I would have to have had a lot more cash beforehand in order to finance my escapades. Some compatibilists (e.g., Saunders, 1968) have argued that incompatibilist defenders of the Consequence Argument rely upon the outlandish notion of ability in the first premise of their argument, but, these compatibilists maintain, the first premise is falsified when interpreted with a milder notion of ability.

### 5.1.2 Challenging Power Necessity and the Laws of Nature

Other compatibilists have argued against the first premise of the Consequence Argument by attempting to show that a person *can* act in such a way that a law of nature would not obtain. As with the distinction drawn regarding ability and the past, consider the difference between a person who has the ability to act in such a way that *she violates a law of nature*, as opposed to a person (at a deterministic world) who has the ability to act in such a way that, *if she were to so act, some law of nature that does obtain would not*. Notice that the former ability would require magical powers. According to the compatibilist, the latter, by contrast, would require nothing outlandish. It merely tells us that a person who acted a certain way at a certain time possessed abilities to act in various sorts of ways. Had she exercised one of those abilities, and thereby acted differently, then the laws of nature that would have entailed what she did in that hypothetical situation would be different from the actual laws of nature that did actually entail what she did actually do. This latter ability does not assume that agents are able to violate laws of nature; it just assumes that whatever the laws of nature are (at least at deterministic worlds), they must be such as to entail, given the past, what an agent will do. If an agent acts differently in some possible world than she acts in the actual world, then some other set of laws will be the ones that entail what she does in that world. Some compatibilists (most notably Lewis, 1981), fixing upon ability pertaining to the laws of nature, have argued that incompatibilist defenders of the Consequence Argument rely upon the outlandish notion of ability in the first premise of their argument. But, these compatibilists maintain, that first premise is falsified when interpreted with an uncontroversial notion of ability.

### 5.1.3 Challenging the Inferences Based upon Power Necessity

Michael Slote (1982) attempted to refute the Consequence Argument by showing that the inference principle upon which the argument relies is invalid. According to Slote, one cannot draw the desired incompatibilist-friendly conclusion even if the Consequence Argument's premises are all true. The central point towards which Slote works is that notions like *unavoidability* (or *power necessity*) are sensitive to contexts in a way that only "selectively" permits the sort of inference at work in the Consequence Argument. Let us work with the idea of unavoidability. According to Slote, when we say that such and such is unavoidable for a person, we have in mind "selective" contexts in which the facts pertaining to the unavoidability have nothing to do with that person — the facts bypass that person's agency altogether (Slote, 1982, p.19). It is unavoidable for me, for instance, that Caesar crossed the Rubicon, or that most motor vehicles now run on gasoline. Nothing about my agency — about what I can do — can alter such facts. This suggests that unavoidability is misapplied when it concerns aspects of a person's own agency. But notice that in the Consequence Argument unavoidability (or power necessity) trades between a context

in which the notion is appropriately applied, and one in which, according to Slote, it is not. In the Consequence Argument, the first premise cites considerations that have nothing to do with a person's agency — facts prior to her birth, and the laws of nature. It is claimed that these facts are unavoidable for a person, but from this a conclusion is drawn that the very actions a person performs are unavoidable for her. And this, Slote and other compatibilists (such as Dennett, 1984a) have suggested, is to draw incompatibilist conclusions illicitly from reasonable claims regarding unavoidability.

#### 5.1.4 Looking Past the Consequence Argument

Suppose that one compatibilist reply or another proves that the Consequence Argument is unsound.<sup>[21]</sup> This alone would not amount to a positive argument for compatibilism. It would merely mean that one argument for the incompatibility of determinism and regulative control is untenable. But that is consistent with the *incompatibility* of determinism and regulative control. Some argue for this incompatibility without relying upon the assumptions at work in the Consequence Argument (Fischer, 1994; and Ginet, 1990, 2003). Furthermore, even if the compatibilist were in a position to discredit all current arguments for the incompatibility of determinism and regulative control, *she would still need a positive argument demonstrating the compatibility of determinism and regulative control*. Otherwise, she would still face the intuitive conflict between a Garden of Forking Paths model of control and the claim of determinism. This conflict is encapsulated in the second premise of the Basic Incompatibilist Argument: If determinism is true, no one can do otherwise than she does (see section 2.1). Hence, supposing the Consequence Argument is defeated, compatibilists wishing to defend regulative control (such as Bernard Berofsky, 1987, 1995) still have their work cut out for them.

### 5.2 Multiple Viewpoints Compatibilism

One influential contemporary defense of compatibilism is Daniel Dennett's. In his 1984 book *Elbow Room*, as well as in several important papers, including “On Giving Libertarians What They Say They Want,” (1981c) and “Mechanism and Responsibility” (1973), Dennett advances compatibilism by drawing upon important developments in the philosophy of mind.

#### 5.2.1 Intentions and Stances

Dennett argues for the legitimacy of folk psychological notions in the explanation of intentional action. His view turns upon the range of stances adopted towards a system, stances that are legitimated by their effectiveness in understanding, predicting, and interacting with the system. According to Dennett, even a thermostat can be interpreted as a very limited intentional system since its behavior can usefully be predicted by attributing to it adequate beliefs and desires to display it as acting rationally within some limited domain. For example, the thermostat *desires* that the room's temperature (or the engine's internal temperature) not go above or below a certain range. If it *believes* that it is out of the requisite range, the thermostat will respond appropriately to *achieve* its desired results.

But surely, it might be objected, a thermostat does not “really” have intentions, not like titmice, toddlers or college freshmen. According to Dennett, this starts one down the wrong path (1973, p. 155).<sup>[22]</sup> To seek a clean distinction between some metaphysically authentic intentional beings from simulacra like thermostats presupposes that there is more to any intentional system than adopting a stance toward it as an intentional system. If that stance genuinely pays off — if it facilitates a fruitful exchange, allows for helpful predictions, allows one to engage rationally with it — then it wins the status of an intentional creature. No special metaphysical tag is needed. Hence, for Dennett, the propriety of adopting the *intentional stance* towards a system is settled pragmatically in terms of the utility of its application in interacting with the system. Along with this thesis goes Dennett's claim that folk psychological explanations (appealing to the intentional stance) are entirely consistent with more basic stances such as the *design* or *physical stances*, the former appealing to the intentions, not of the system, but of its designer, the latter appealing only to the basic mechanistic processes that cause the system from moment to moment to move from one physical

state into another. Once a system becomes sufficiently complex, as with even a chess playing computer, the intentional stance will become indispensable for successful interaction (1973, p. 154).

### 5.2.2 The Intentional Stance and the Personal Stance

Dennett makes use of his treatment of the intentional stance to argue for compatibilism. Just as the decision to adopt towards a system the intentional stance is a pragmatic one, so too is it a pragmatic decision to adopt towards a system the stance that it is a morally responsible person. Dennett calls this latter stance the *personal stance* (1973, pp. 157-8). As with the intentional stance, there is nothing metaphysically deep required to interpret legitimately a system as a person (no special faculty of the will for instance). Such systems are morally responsible agents if interpreting them according to the personal stance pays off (1984a, pp. 158-63). And of course, just as the physical (or the deterministic) stance is compatible with the intentional stance, so too, according to Dennett, is it compatible with the personal stance. Furthermore, just as he treats the intentional stance, Dennett argues that, due to the complexity of such systems, it is practically impossible to interpret and predict the system purely from the physical (deterministic) stance. Hence, the physical stance will never supplant the personal stance. We persons involved in the everyday commerce of interacting with each other need the personal stance; it is not threatened by the specter of determinism. Let us call Dennett's view, *Multiple Viewpoints Compatibilism*.

### 5.2.3 Dennettian Free Will

What is free will on Dennett's account? Dennett explicitly rejects regulative control (1984a, 1984b), arguing for a point that he shares with Frankfurt (1969), namely, that the ability or inability to do otherwise is irrelevant to the control pertinent to moral responsibility. But how does Dennett account for guidance control? For Dennett, free will consists in the ability of a person to control her conduct on the basis of rational considerations through means that arise from, or are subject to, critical self-evaluation, self-adjusting and self-monitoring. That is, free will involves *responsiveness to reason*. Dennett certainly has many useful observations about how this sort of control might have naturally arisen from less sophisticated sorts of creatures through a process of evolution (1981b). Later on other philosophers offered careful explications of control understood as such. (See John Martin Fischer's work, discussed below in section 5.5.)

### 5.2.4 Dennett versus the Source Incompatibilist

But what about the Source model of control, as well as the Source Incompatibilist Argument (section 2.2)? How does Dennett's multiple viewpoints compatibilism stack up against them? Against the crucial first premise of the Source Incompatibilist Argument — A person acts of her own free will only if she is its ultimate source — in his book *Elbow Room*, Dennett, it seems, wants to place the incompatibilist on the defensive, arguing that it is only confusion driven by appeal to what he calls “intuition pumps” that makes the premise seem at all plausible. Intuition pumps, according to Dennett, are examples designed to sway our philosophical intuitions, but are themselves philosophically suspect. Such examples involve cases of (apparently) normally functioning agents being manipulated, for example, as if like a puppet hooked up to some wires. But Dennett's polemical approach might seem dialectically unfair. Not all worries about the sources of action are groundless, even those arising from “intuition pumps” built on cases of ghastly manipulation. Dennett's incompatibilist opponent deserves more credit than he seems willing to give her. Regardless of his dismissive attitude towards the notion of ultimacy, and of an argument like the Source Incompatibilist Argument, Dennett's positive account of morally responsible agency certainly does take very seriously a source model of free will. By appealing to views on intentionality, rational action, agency, and personhood, Dennett offers a suggestive account of how it is that an agent can be an authentic source of her action (1984a, pp. 50-73).

## 5.3 Hierarchical Compatibilism

Perhaps the most widely recognized form of contemporary compatibilism is Harry Frankfurt's hierarchical mesh theory (1971). Frankfurt's theory explains freely willed action in terms of actions that issue from desires of a certain sort. In particular, a desire issuing in a freely willed action must suitably mesh within hierarchically ordered elements of a person's psychology. The key idea is that a person who acts of her own free will acts from desires that are nested within more encompassing elements of her self. Hence, Frankfurt develops a source model of control to explain how it is that, when a freely willing agent acts, her actions emanate from *her* rather than from something foreign.<sup>[23]</sup>

### 5.3.1 Higher-Ordered Desires and the Nature of Persons

Frankfurt distinguishes between first and second-order desires. This serves as the basis for his hierarchical account. The former desires have as their objects actions, such as eating a slice of cheese cake, taking in a movie, or gyrating one's hips to the sweet sounds of B. B. King. The latter are desires about desires. They have as their objects, desires of the first-order, such as the desire to have the motivation to exercise daily (something that, regrettably too many of us lack): "If only I *wanted* to go to the gym today, then it would be easy for me to get my tail off this couch!"

Amongst the first-order desires that a person has, some are ones that *do not* move her to action, such as one's (unsatisfied) desire to say to her boss what she knows that she should not. Other first-order desires, however, *do* move a person to action, such as one's (satisfied) desire to follow through on her boss's request. Frankfurt identifies an agent's *will* with her effective first-order desire, the one moving a person, as Frankfurt puts it, "all the way to action" (1971, p.84).<sup>[24]</sup>

Frankfurt also distinguishes between different sorts of second-order desires. Some are merely desires to have first-order desires, but *not* first-order desires that would comprise her will. Frankfurt uses the example of a psychotherapist who wishes to experience a desire for narcotics so as to understand a patient better. The therapist has no wish that this desire be effective in leading her to action (1971, pp.84-5). She wants to know what it is like to feel the craving for the drug; she has no wish to take it. On the other hand, other second-order desires that a person has are desires for *effective* first-order desires, desires that would comprise her will, and would thereby be effective in moving her all the way to action. For instance, the dieter who is constantly frustrated by her sugar cravings might desire a more effective desire for health, one that would be more effective in guiding her eating habits than it often is. These second-order desires Frankfurt calls *second-order volitions*. There is no theoretical limit to how highly-ordered one's desires might be. The dieter in the above example might develop a third-order desire for her second-order desire (regarding her desire for health) not to play such a dominant role in her daily deliberations. Other things, she might reason, are of more importance in life than concerning herself with her dietary motivations.

According to Frankfurt, a distinctive feature of personhood is that only a person has second-order volitions. Only a person wants to be moved by different desires and motives from the ones that move her. Frankfurt calls agents who have no second-order volitions *wantons*. Persons care about which desires lead them to action. Wantons do not. They are passive bystanders to their wills (1971, p. 89).

### 5.3.2 Frankfurt's Hierarchical Theory of Free Will

Frankfurt uses the examples of three different sorts of addicts to illustrate his concept of free will. Consider first the *wanton addict*. She has conflicting first-order desires. She desires both to take the drug to which she is addicted, as well not to take the drug. But the wanton addict has no higher-order volition regarding which of her first-order desires wins out. She is passive with regard to the battle of desires taking place within her. The *unwilling addict*, like the wanton addict, has both a first-order desire to take the drug, and a first-order desire not to take the drug. But, unlike the wanton addict, the unwilling addict also has a second-order volition that her first-order desire to take the drug *not be her will*. This is the basis for her unwillingness. Regrettably, her irresistible addictive desire to take the drug constitutes her will. Finally, consider the case of the *willing addict*. The willing addict, like both the wanton and the unwilling addict, has conflicting first-order desires as regards taking the drug to which she is addicted. But the willing addict,

by way of a second-order volition, embraces her addictive first-order desire to take the drug. She wants to be as she is and act as she does.

It is now easy to illustrate Frankfurt's hierarchical theory of free will. The wanton is not a person, and so is not a candidate for freely willed action. The unwilling addict does not take the drug of her own free will since her will conflicts at a higher level with what she wishes it to be. The willing addict, however, takes the drug of her own free will since her will meshes with what she wishes it to be. Frankfurt's theory can now be set out as follows:

One acts of her own free will if and only if her action issues from the will she wants

It might seem strange that Frankfurt's willing addict acts of her own free will since, due to her addiction, she could not do otherwise. But recall (sections 4.2, and 5.1), Frankfurt does not believe that freedom involving alternative possibilities is required for moral responsibility. Frankfurt instead believes that the freedom pertinent to moral responsibility concerns what an agent does do and her actual basis for doing it. That is, Frankfurt believes that it is guidance control that is necessary for moral responsibility, not regulative control. The willing addict possesses the sort of freedom required for moral responsibility because the will leading to her action is the one that she wishes it to be; she acts with guidance control.

Frankfurt's theory has been categorized as a *Real Self Theory* (Wolf, 1990, p. 29). It is easy to see why. According to Frankfurt, the sort of freedom needed for assessments of moral responsibility turns crucially on whether or not the agent reveals herself in acting as she does, or if instead her conduct is in some way alien to her. By desiring at a hierarchically higher-order level of reflection that one's will be a certain way (or not be a certain way), one reveals her deeper self, not merely at the surface of her conduct, but in terms of how she herself regards her very own motives issuing in her conduct. When she acts of her own free will, those motives are hers, are *of* her. She owns them. Hence, they reflect her true self. When she acts, but does not act of her own free will, she disavows her motives. They do *not* reflect her true self.

### 5.3.3 Two Problems for a Hierarchical Theory

Frankfurt's hierarchical theory has been under intense scrutiny since he first presented his position. We shall consider here two objections that emerge from structural aspects of it. One has to do with its hierarchical nature. The other has to do with its relying exclusively upon a mesh between different features of an agent's psychology. (For a discussion of Frankfurt's attempts to respond to these problems, see section 6 below.)

Consider the *hierarchical problem*. According to Frankfurt, a person facing a problem with regard to her will's freedom faces a situation in which her first-order desires are in conflict. On Frankfurt's theory, a person has the resources to form second-order desires as to which of her conflicting first-order desires should move her. By this means, an agent endorses one of the first-order desires and, if all goes smoothly, that one becomes her will. Through this process, she draws within the sphere of her self one sort of desire and alienates another. Now, the problem is simply this: If a person can be conflicted at the level of her first-order desires, she can also be conflicted at the second, or even at higher-orders (Watson, 1975). Hence, the problem of an agent's free will can reappear at these ever ascending stages. If this is correct, Frankfurt's view is incomplete. Maybe his account of free will does articulate a necessary condition for acting of one's own free will, but it appears not to be sufficient. It needs supplementing so as to avoid the problem of a spiraling reoccurrence of challenges to an agent's freedom.

Next consider the *mesh problem*. According to Frankfurt, if freely willed action for which an agent is morally responsible is purely a function of the relation between an agent's will and her second-order volitions, then Frankfurt is committed to the view that it does not matter in any way how an agent came to have that particular mesh. But cases can be constructed that seem to suggest that it *does* matter how an agent came to have the particular mesh between her first and second-order desires. (For example, see Slote, 1985; and Fischer and Ravizza, 1998, pp.194-206). Using Frankfurt's own example of the willing addict,

suppose that the addict's second-order willingness is itself caused by the effects of the drug use. Suppose that the drug use has impaired her evaluations or preferences arising at a second-order of reflection on her own mental states. Or, setting this sort of case aside, imagine that an agent is brainwashed or manipulated through some means or another, say by hypnosis, or by aliens zapping a person into having a different set of psychological preferences than those that she would otherwise have. In all of these cases, just call them *manipulation cases*, Frankfurt seems committed to the view that such agents act of their own free will and are morally responsible so long as the appropriate psychological mesh is in place, no matter what sort of (merely apparent) freedom and responsibility-undermining history gave way to an agent's having that particular mesh.

### 5.3.4 Frankfurt versus the Source Incompatibilist

Grant that Frankfurt is correct that free will and moral responsibility do not require regulative control. How does Frankfurt's view stack up against the Source Incompatibilists? Frankfurt develops an account of free will out of a perceptive treatment of what it is to be a person and how it is that a person's will permits a level of depth, of self-awareness and reflection, that can emerge in a person's conduct. Frankfurt's is a rich Source model of agency, one carrying moral depth. For Frankfurt, an agent that acts of her own free will does not merely reveal her desires in action, she reveals how she wishes herself to be as a person.

How might Frankfurt reply to the Source Incompatibilist Argument? Naturally, he must resist the first premise — a person acts of her own free will only if she is its ultimate source (see section 2.2). But how? Recall that an agent is an ultimate source of her action only if no conditions external to her are sufficient for her action. Determinism clearly is incompatible with this. Frankfurt's battle with the source incompatibilist must turn on showing that his account of the source of a person's free agency is sufficient; ultimacy is not needed. But now, consider the manipulation cases that challenged Frankfurt's reliance exclusively upon a mesh between different constituents in an agent's psychology. To the extent that the manipulation cases suggest that the mesh can arise in a freedom and responsibility-undermining way, it seems that Frankfurt's treatment of the proper source of freely willed action is incomplete. Frankfurt needs to show what is defective in a mesh being brought about in these deviant manners (and how mere determination does not share these defective features of the manipulation cases). If he cannot, then it seems that the source incompatibilist has the upper hand. She can say that, if one sort of causal history giving rise to a Frankfurtian mesh can undermine an agent's freedom and moral responsibility, then why not a deterministic history? A deterministic history is simply a more elaborate form of manipulation that happened to take a very long time to achieve the same sort of result. To join issue with the source incompatibilist, Frankfurt must either show why manipulation cases fail, or instead, must bite the bullet and accept that, on his theory, agents so manipulated can still be free and morally responsible persons.

### 5.4 The Reason View

In her artfully crafted book, *Freedom within Reason* (1990), as well as in several provocative papers (1980, 1987), Susan Wolf develops a mesh theory between an agent's actions and values. For Wolf, free will concerns an agent's ability to act in accord with the True and the Good. Because the conditions of Wolf's mesh theory require an anchor external to the agent's internal psychological states (the True and the Good), unlike Frankfurt's, hers is not a real self theory (1990, pp.73-76) The crucial question is not just whether an agent reveals her deeper self in her conduct (1987); it is whether she is able to act upon moral reason. Hence, Wolf embraces the title, *The Reason View*.

In her effort to make free will track moral reason, Wolf develops an asymmetry thesis according to which praiseworthy conduct does not require the freedom to do otherwise but blameworthy behavior does (1980; and 1990, pp.79-81). Put in terms of guidance and regulative control, only blameworthy conduct requires regulative control. Guidance control is sufficient for praiseworthy conduct. Wolf's reasoning is that, if an agent does act in accord with the True and the Good, and if indeed she is so psychologically determined that she cannot but act in accord with the True and the Good, her inability to act otherwise does not threaten the sort of freedom that morally responsible agents need. But blameworthy behavior, Wolf reasons, does require regulative control since, if an agent acts contrary to the True and the Good, but is so

psychologically determined that she cannot act in accord with it, then, being unable to act as reason requires, it would be unreasonable to blame her.

#### 5.4.1 Wolf's Asymmetry Thesis and Frankfurt Examples

Wolf's view differs from Frankfurt's since hers requires regulative control somewhere. Hence, her compatibilism is open to refutation by incompatibilist arguments designed to show that determinism is incompatible with freedom involving alternative possibilities.<sup>[25]</sup> But it is unclear that Wolf should be committed to the asymmetry. Could she preserve the central feature of her Reason View without requiring regulative control as a condition of blameworthiness? Frankfurt-type examples (see section 4.2) can be constructed for cases of blameworthy action that seem to suggest that Wolf should give up the requirement of regulative control for blameworthiness. (See Fischer and Ravizza, 1998.)

#### 5.4.2 Assessing Wolf's Reason View

How might Wolf face the two models of control discussed above, as well as the two related incompatibilist arguments? If she wishes to preserve her asymmetry thesis, she must retain some sort of Garden of Forking Paths model for the control required of blameworthy conduct. In this case, she will need to address the crucial premise in the Basic Incompatibilist Argument that holds that an agent cannot do otherwise if determinism is true (see section 2.1). This premise, supported as it is by the Consequence Argument and near cousins of the Consequence Argument (see section 4.1), will demand of Wolf that, minimally, she show what is wrong with arguments like the Consequence Argument, and optimally, that she offer some positive compatibilist account of the ability to do otherwise.

Setting aside questions regarding Garden of Forking Paths freedom, how well does Wolf's Reason View jibe with the Source model of control, as well as the Source Incompatibilist Argument? On Wolf's view, if an agent does act from reasons, and if her reasons are (or are susceptible to) the True and the Good, then she as an agent is a source of conduct that carries with it (or is able to carry with it) the stamp of moral reason. Enough said. But what about the Source Incompatibilist Argument, and the premise concerning ultimacy that seems to plague most every brand of compatibilism: A person acts of her own free will only if she is its ultimate source (see section 2.2)? Like Frankfurt's mesh theory, Wolf's too is endangered by the thought that an agent could be artificially manipulated in a responsibility-undermining manner into satisfying the mesh Wolf's theory demands. And mightn't such manipulation be no different than the manner in which a deterministic world shapes an agent to have the psychological structure and motives she has? Does not the prospect of manipulation cases show that without ultimacy, an agent cannot be the proper source of her action? So it appears that Wolf is at the same crossroads as is Frankfurt. Either she must show what is defective in the manipulation cases so as to distinguish agents so manipulated from the sort of proper mesh demanded by her theory, or she must bite the same bullet and accept that these sorts of manipulated agents, by the conditions of her theory, do act of their own free wills and are morally responsible for their conduct.

### 5.5 Reasons-Responsive Compatibilism

Several compatibilists have suggested that freely willed actions issue from volitional features of agency that are sensitive to an appropriate range of reasons (e.g., Dennett, 1984a; Fingarette, 1972; Gert and Duggan, 1979; Glover, 1970; MacIntyre, 1957; Neely, 1974; and Nozick, 1981). Such reasons might speak in varying ways for or against a course of action. Agents who are unresponsive to appropriate rational considerations (such as compulsives or neurotics) do not act of their own free will. But agents who *are* responsive to some range of rational considerations do. This view has been artfully refined in recent years by John Martin Fischer (1987, and 1994), and subsequently, Fischer and Mark Ravizza (1998). (For a more advanced discussion of Fischer and Ravizza's view, see section 6.) Many working on the topics of free will and moral responsibility now regard Fischer's developed account to be the gold standard for cutting edge defenses of compatibilism.

A *reasons-responsiveness theory* turns upon dispositional features of an agent's relation to reasons issuing in freely willed action. Appropriately reasons-responsive conduct is sensitive to rational considerations. The view is not merely that an agent would display herself in some counterfactual situations to be responsive to reasons, but rather that her responsiveness to reasons in some counterfactual situations is evidence that her actual conduct *itself* — the causes giving rise to *it* — is also in response to rational considerations. (Amendments need to be added to accommodate cases of spur-of-the moment, or impulsive freely willed action).

### 5.5.1 Agent-Based Reasons-Responsiveness

The most natural way to understand a reasons-responsive theory is in terms of an *agent's* responsiveness to reasons. To illustrate, suppose that Frank Zappa plays the banjo of his own free will. According to a reasons-responsive theory, his playing the banjo freely at that time requires that if, in at least some hypothetical cases, he had reason not to, then he would refrain from playing the banjo. For instance, if Jimmy Hendrix were to have stepped into Frank's recording studio and asked Frank to play his electric guitar, Frank would have wanted to make Jimmy happy and thus would have gladly put his banjo aside and picked up his electric guitar. It seems, then, that for Frank to play the banjo of his own free will, Frank — *the agent* — must have regulative control and not merely guidance control over his playing. His freedom must consist partially in his ability to act upon alternatives.

### 5.5.2 A Tension Between Reasons-Responsiveness and Frankfurt Examples

Notice that, because Frankfurt examples challenge the incompatibilists' demand for regulative control, they also challenge an *agent-based* reasons-responsive theory (Fischer and Ravizza, 1998, pp. 34-41). For imagine that the benevolent demon Jerry Garcia wants Frank to play the banjo at the relevant time. Jerry would much prefer that Frank play the banjo on his own. But worried that Frank might elect not to play the banjo, Jerry covertly arranges things so as to manipulate Frank if the need arises. If Frank should show any indication that he will not play the banjo, Jerry will manipulate Frank so that Frank will play the banjo. Hence, when Frank does play the banjo uninfluenced by Jerry's possible intervention, he does so of his own free will. But he has neither regulative control, nor does he seem to be reasons-responsive with respect to his banjo playing. Due to Jerry's presence, he cannot but play the banjo *even if Jimmy Hendrix were to ask Frank to play his guitar*.

To alleviate the tension between a reasons-responsive theory and Frankfurt examples, Fischer argued that reasons-responsive compatibilism can be cast in such a way that it involves only guidance control. Consider the example with Frank, Jimmy, and Jerry. Frank did not have regulative control over his playing the banjo since Jerry's presence insured that Frank play the banjo even if Jimmy were to ask Frank to play his guitar. The scenario in which Jimmy asks Frank not to play his banjo is one that Frank normally *would* find to be a compelling reason to refrain from his banjo playing. Hence, *by his own lights*, Frank *would* find Jimmy's request compelling. Yet, due to Jerry's presence, *Frank* is not responsive to such a weighty reason. What would be required to illustrate responsiveness would be to subtract Jerry from the scenario. This would do the trick. So suppose that Frank plays the banjo of his own free will, even with Jerry passively standing by. How can it be shown that Frank's conduct was, in some manner, reasons-responsive? How can it be shown that what he actually did was in response to a reason? Well, *if* Jimmy Hendrix had asked Frank not to play the banjo but the guitar instead, *and* if Jerry's presence were to be subtracted from the situation, *then* Frank would respond to Jimmy's request and play the guitar and not the banjo. This shows that Frank does play the banjo of his own free will even in the actual situation in which Jerry is passively standing by.

### 5.5.3 A Mechanism-based Reasons-Responsive Theory

Illustrating reasons-responsiveness in a Frankfurt example *does* require recognizing counterfactual conditions in which an agent acts otherwise in response to reasons. But in a Frankfurt example, one has to subtract from those conditions the presence of the insuring conditions (the demon) designed to guarantee that the agent not act otherwise. How can this move be legitimate? How is it not just an arbitrary addendum

to cram together two compatibilist themes that otherwise appear to be at odds (reasons-responsiveness and Frankfurt examples)? It is not arbitrary, and here is why. Think about what happens in the actual scenario of a Frankfurt example. As things unfold, the demon is *inactive*. The agent acts for her own reasons. But now, focusing solely on *what the agent does* in this actual scenario, and the reasons that give her a basis for doing what she does, consider what deliberative features of her agency played the casual role in the *actual sequence* of events bringing about her action. To capture what features of deliberative agency do play a role in the actual causal sequence of an agent's action, not every element of the agent seems to be involved in the process. For instance, in the example above, in playing the banjo of his own free will, Frank might have a large range of beliefs and desires that is entirely irrelevant to the range that did play a causal role in his action. Frank might have believed that chickens are both feathered and not toasters, and he might have desired that his toenails be painted purple. But neither of those elements of his psychological state needed to play a role in the sorts of factors that did lead him to play the banjo. So, whatever that narrower range of agential characteristics within the wider spectrum of all of the features that made up Frank Zappa, the agent, just fix on *that* narrow spectrum. Since it is just that narrow spectrum that we shall now identify with the causal production of Frank's conduct, just call it the *mechanism* of his action.

Once we have located the mechanism of action that is at work in the actual causal sequence of a Frankfurt example, we can turn our attention to understanding the dispositional features of *it* as a casual mechanism. If other reasons bear upon it, then it would be sensitive to some of those reason. It would produce different conduct in some reasonable range of cases. If it would, *then that very mechanism is responsive to reasons*. Confirming that that very mechanism is responsive to reasons would not merely illustrate that, in scenarios other than the actual, the agent acts upon a mechanism sensitive to reason. It would also illustrate that in the Frankfurt scenario in which the agent really *does* act, what *does* play a role in the actual causal sequence of her action is some feature of *her* agency (a mechanism) that *itself* is *in fact* a response to a reason.

Fischer offers an *actual-sequence, mechanism-based, reasons-responsive* analysis of guidance control. He maintains that his analysis of guidance control is compatible with determinism. According to Fischer, an agent, and the mechanism of her action, can be entirely determined in the actual sequence of events in which she acts. Yet the actual manner in which her mechanism responds to reasons could be appropriately sensitive to reasons such that, if different reasons were to bear upon it, it would respond differently, and the agent whose mechanism it is would act differently than she does act.

#### 5.5.4 Assessing Reasons-Responsive Compatibilism

How might a reasons-responsive approach compare with a view like Frankfurt's hierarchical model? To demarcate the relevant wants issuing in freely willed conduct, Frankfurt needed to postulate a higher-order of willing into which effective desires meshed. Only then could Frankfurt help compatibilists distinguish freedom-undermining wants (such as those involved in compulsive conduct) from freedom-conferring wants. Reasons-responsiveness attempts to fix this sort of problem by different means. Instead of postulating a hierarchical relation between different sorts of wants or desires, it instead gives a dispositional analysis of the wants' or desires' (the reasons') sensitivity to rational considerations. The differences might well yield different ways of treating some cases. For example, a willing addict would not be reasons-responsive and so would not take the drug of her own free will according to a reasons-responsive compatibilist. But according to Frankfurt, a willing addict does take the drug of her own free will.

How might a mechanism-based reasons-responsive theory satisfy a source model of control? A view such as Fischer's *might* be at a disadvantage. It is possible that a reasons-responsive mechanism could be unhitched from the agent whom it affects. So it is an open question whether or not Fischer's compatibilist position offers as rich an account of the source of an agent's action as does a real self view such as Frankfurt's. Of course, if Fischer is able to advance an ownership condition that does anchor an agent's reasons-responsive mechanism to the agent's self, then Fischer's view will not only do the work that Frankfurt's does in accommodating a source model of control (linking a freely willed action to a real self), it will also do the sort of work Wolf's does. That is, Fischer's view will then show how, as a source of her conduct, a morally responsible agent can be tightly linked to reasons (reasons Wolf identifies under the heading, "the True and the Good").

How does Fischer's view stack up against the Source Incompatibilist Argument? The challenge Fischer faces here is the same faced by Frankfurt and Wolf. The source incompatibilist maintains that it is a necessary condition of free will that one be an ultimate source of her action, and determinism is incompatible with one's being an ultimate source of her action (see section 2.2). The compatibilist's task is to show that her treatment of the source of an agent's conduct is sufficient for free will. But the source incompatibilist will point to manipulation cases that suggest that some causal histories giving rise to compatibilist friendly psychological structures, such as reasons-responsive mechanisms, are freedom and responsibility-undermining. If so, then why is determinism any different from a manipulation case? The burden, it seems, is upon the compatibilist to distinguish how it is that manipulation cases differ from a normal deterministic history. The compatibilist's only other strategy is simply to deny that the pertinent manipulated agents are not free and morally responsible. (For Fischer's efforts to avoid the problem, see section 6.)

## 5.6 Strawsonian Compatibilism

It would simply be misleading not to mention Strawsonian compatibilism amongst the views characterizing contemporary compatibilism. The secondary literature devoted to it proves that it is still alive and well in contemporary debate.

Several contemporary philosophers have advanced Strawsonian themes. For instance, John Martin Fischer and Mark Ravizza hold that an account of guidance control aids in providing the conditions of application for the concept of moral responsibility, a concept that they maintain is Strawsonian (1998, pp.1-27). Fischer and Ravizza intend their Strawsonian theory as an amendment to Strawson's suggestion that moral responsibility is to be associated with the reactions of those within the moral community to members of the community. Fischer and Ravizza advise that moral responsibility be developed by thinking in terms of the *propriety conditions* for the morally reactive attitudes. This amendment would have it that a moral community could respond to a group of persons inappropriately, either failing to recognize persons who are free and morally responsible agents (slaves, for example), or instead including beings who are not (for instance, very young children, farm animals, or the weather).

Gary Watson also embraces Strawsonian compatibilism (1987). Watson sought to elaborate upon it by thinking of our moral responsibility practices, and in particular the morally reactive attitudes, along the lines of a communication-based theory in which a morally responsible agent's competence turns in some way upon being a potential interlocutor to moral conversations between her and the moral community in which she operates. On this view, the control condition for moral responsibility would have to fit the capacity to communicate morally through word and deed with members of the moral community.

Susan Wolf defends (with significant reservations) the Strawsonian thesis that the interpersonal viewpoint (that permits access to the morally reactive attitudes) is one that a freely willing agent cannot give up (1981). Wolf diverges at points with Strawson's own manner of defending this. But Wolf's central thesis is Strawsonian. A person cannot fully forswear the point of view of the interpersonal attitudes, and this point of view is the point of view from whence our morally reactive attitudes gain their force and figure in our conduct.

Paul Russell (1995) has also defended a form of Strawsonian compatibilism, the central features of which he finds anticipated in Hume's writings on free will and moral responsibility. According to Russell, we can learn from Hume, as Strawson did, to understand our moral responsibility practices as fundamentally a matter of our sentiments and our social expectations as structured and sustained by these sentiments. Fixing on our moral natures, as we should, dispels any presumption that determinism would somehow pose a threat to our conceptions of freedom and moral responsibility.

One pressing question for Strawsonian compatibilism is how much emphasis should be placed upon the point of view of those in the moral community who hold others morally responsible. On a strong, and radically anti-metaphysical reading, those in the moral community determine the conditions for when a

person is or is not a morally responsible agent, as well as whether a person is or is not morally responsible for some bit of conduct. On this view, morally responsible agency is to be extrapolated from the practice of the members of the moral community in holding persons morally responsible. This suggests a compatibilist strategy according to which the freedom required for moral responsibility derives from the normative considerations embraced by the members of the community holding persons responsible. In its strongest form, according to this sort of compatibilist approach, there need be *no* threat to freedom or moral responsibility from determinism since a community can *construct* a set of standards for freedom and responsibility that could be satisfied even in a determined world. Given that the conditions are constructed, they need not be constrained by prior metaphysical questions concerning the nature of the persons alleged to possess free will. The community will, so to speak, settle matters of what free will is, not the underlying nature of the person whose free will is at issue. This theme, suggested in Strawson's famous 1962 essay, is developed by Jay Wallace in *Responsibility and the Moral Sentiments* (1994). Wallace's position has emerged as a serious alternative to the sorts of approaches to the free will problem that take as their theoretical starting point the nature of the persons, or the action-theoretic characteristics of the process issuing in freely willed action.

These issues are pursued further in the supplement:

### [Compatibilism: The State of the Art](#)

#### **Bibliography**

- Ayer, A. J. 1954. 'Freedom and Necessity.' In his *Philosophical Essays*. New York: St. Martin's Press: 3-20; reprinted in Watson, ed., 1982, pp. 15-23.
- \_\_\_\_\_. 1980. 'Free-Will and Rationality.' In van Straaten, ed. 1980.
- Bennett, Jonathan. 1980. 'Accountability.' In van Straaten, ed. 1980.
- Berofsky, Bernard. 2002. 'Ifs, Cans, and Free Will: The Issues.' In Kane, ed., 2002.
- \_\_\_\_\_. 1995. *Liberation from Self*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. 1987. *Freedom from Necessity*. London: Routledge & Kegan Paul.
- \_\_\_\_\_. ed., 1966. *Free Will and Determinism*. New York: Harper & Row.
- Bok, Hilary. 1998. *Freedom and Responsibility*.
- Buss, Sarah, and Lee Overton. 2002. *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge, Mass: MIT Press.
- Chisholm, Roderick. 1964. 'Human Freedom and the Self.' *The Lindley Lectures*. Copyright by the Department of Philosophy, University of Kansas. Reprinted in Watson, ed., 1982.
- Clarke, Randolph. 2000. 'Incompatibilist (Nondeterministic) Theories of Free Will', *The Stanford Encyclopedia of Philosophy* (Fall 2000 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/fall2000/entries/incompatibilism-theories/>.
- Clarke, Randolph. 1993. 'Toward a Credible Agent-Casual Account of Free Will.' *Noûs* 27, pp.191-203.
- Dennett, Daniel. 1984a. *Elbow Room: Varieties of Free Will Worth Wanting*. Cambridge, Mass: MIT Press.
- \_\_\_\_\_. 1984b. 'I Could Not Have Done Otherwise — So What?' *The Journal of Philosophy* LXXXI, 10: 553-67.
- \_\_\_\_\_. 1981a. *Brainstorms: Philosophical Essay on Mind and Psychology*. Cambridge, Mass.: MIT Press, 1981), pp. 286-99.
- \_\_\_\_\_. 1981b. 'Conditions of Personhood.' In Dennett, 1981a: 267-86.
- \_\_\_\_\_. 1981c. 'On Giving Libertarians What They Say They Want.' In Dennett, 1981a: 286-99.
- \_\_\_\_\_. 1973. 'Mechanism and Responsibility.' In Lehrer (1973). Reprinted in Watson, ed. 1982.
- Eshleman, Andrew S. 2001. 'Being is not Believing: Fischer and Ravizza on Taking Responsibility.' *Australasian Journal of Philosophy* 79: 479-90.
- Fingarette, Herbert. 1972. *The Meaning of Criminal Insanity*. Berkeley: University of California Press.
- Fischer, John Martin. 1999. 'Recent Work on Moral Responsibility.' *Ethics* 110: 93-139.

- -----1994. *The Metaphysics of Free Will*. Oxford: Blackwell Publishers.
- -----, 1987. 'Responsiveness and Moral Responsibility.' In Schoeman 1987: 81-106.
- -----, ed., 1986. *Moral Responsibility*. Ithaca: Cornell University Press.
- -----, 1986. 'Power Necessity.' *Philosophical Topics* 14: 77-91.
- -----, 1983. 'Incompatibilism.' *Philosophical Studies*. 43: 127-37.
- -----, 1982. 'Responsibility and Control.' *Journal of Philosophy* 89: 24-40.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.
- -----, eds. 1993. *Perspectives on Moral Responsibility*. Ithaca: Cornell University Press.
- Frankfurt, Harry. 2002. 'Reply to John Martin Fischer.' In Buss and Overton, eds., 2002.
- -----, 1999. *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.
- -----, 1994. 'Autonomy, Necessity, and Love.' In Fulda and Horstman, eds. 1994. Reprinted in Frankfurt, 1999.
- -----, 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- -----, 1987. 'Identification and Wholeheartedness.' In Schoeman, ed. 1987. Reprinted in Frankfurt, 1999.
- -----, 1971. 'Freedom of the Will and the Concept of a Person.' *Journal of Philosophy* 68: 5-20. Reprinted in Fischer, ed., 1986; Frankfurt, 1987; and Watson, 1982.
- -----, 1969. 'Alternate Possibilities and Moral Responsibility.' *Journal of Philosophy* 66: 829-39. Reprinted in Fischer 1986 and Frankfurt 1987.
- Fulda, H.F., and R.-P. Horstmann, eds. 1994. *Vernunftbegriffe in der Moderne: Stuttgart Hegel-Kongress 1993*. Stuttgart: Klett-Cotta.
- Gert, Bernard, and Tim Duggan. 1979. 'Free Will as the Ability to Will,' *Nous* 13: 197-217. Reprinted in Fischer, 1986.
- Ginet, Carl. 2003. 'Libertarianism.' In Loux and Zimmerman, eds., 2003.
- -----, 1996. 'In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing.' *Philosophical Perspectives* 10: 403-17.
- -----, 1990. *On Action*. Cambridge: Cambridge University Press.
- -----, 1983. 'In Defense of Incompatibilism.' *Philosophical Studies* 44, 391-400.
- -----, 1980. 'The Conditional Analysis of Freedom.' In van Inwagen, ed., 1980.
- -----, 1966. 'Might We Have No Choice?' In Lehrer, 1966: 87-104.
- Glover, Jonathan. 1970. *Responsibility*. New York: Humanities Press.
- Haji, Ishtiyaque. 2002. 'Compatibilist Views of Freedom and Responsibility.' In Kane, ed., 2002.
- -----, 1998. *Moral Appraisability*. New York: Oxford University Press.
- Hobart, R. E. 1934. 'Free Will as Involving Indeterminism and Inconceivable Without It.' *Mind* 43: 1-27.
- Hobbes, Thomas. 1997. *Leviathan*. R.E. Flatman & D. Johnston, eds. New York: W.W. Norton & Co.
- Honderich, Ted. 1988. *A Theory of Determinism*. Oxford: Clarendon Press.
- -----, ed. 1973. *Essays on Freedom and Action*. London: Routledge & Kegan Paul.
- Howard-Snyder, Daniel, and Jeff Jordan, eds., 1996. *Faith, Freedom, and Rationality*. Lanham, MD: Rowman and Littlefield.
- Hume, David. 1978. *An Enquiry Concerning Human Understanding*. Ed., P.H. Niditch. Oxford: Clarendon Press.
- -----, 1978. *A Treatise of Human Nature*. Ed., P.H. Niditch. Oxford: Clarendon Press.
- Hunt, David P. 2000. 'Moral Responsibility and Unavoidable Action.' *Philosophical Studies* 97: 195-227.
- Kane, Robert, ed. 2002. *The Oxford Handbook of Free Will*. Oxford and New York: Oxford University Press.
- -----, 1996. *The Significance of Free Will*. Oxford: Oxford University Press.
- Kapitan, Tomis. 2002. 'A Master Argument for Incompatibilism?' In *The Oxford Handbook of Free Will*. ed. Robert Kane. New York: Oxford University Press, pp. 127-57.
- Lamb, James. 1977. 'On a Proof of Incompatibilism.' *Philosophical Review*. 86: 20-35.

- Lehrer, Keith, ed. 1966. *Freedom and Determinism*. New York: Random House.
- Lewis, David. 1981. 'Are We Free to Break the Laws?' *Theoria* 47: 113-21.
- Loux, Michael and Dean Zimmerman, eds., 2003. *Oxford Handbook of Metaphysics*. Oxford: Oxford University Press.
- MacIntyre, Alisdair. 1957. 'Determinism.' *Mind*. 66: 28-41.
- McKay, Thomas and David Johnson. 1996. 'A Reconsideration of an Argument against Incompatibilism.' *Philosophical Topics* 24: 113-22.
- McKenna, Michael. 2003. 'Robustness, Control, and the Demand for Morally Significant Alternatives.' In *Moral Responsibility and Alternative Possibilities*. Eds., Widerker and McKenna. 2002.
- ----. 2002. 'Hilary Bok's, *Freedom & Responsibility*.' Review. *Ethics*. Vol. 113, No. 1: 144-5.
- ----. 2001a. 'Ishtiyaque Haji's, *Moral Appraisability*.' Critical notice. *Philosophy and Phenomenological Research* 63: 711-5.
- ----. 2001b. 'John Martin Fischer and Mark Ravizza's *Responsibility & Control*.' Review. *Journal of Philosophy*. XCVIII, No. 2: 93-100.
- ----. 2000. 'Assessing Reasons-Responsive Compatibilism.' *International Journal of Philosophical Studies*. Vol. 8, No.1: 89-114.
- ----. 1998. 'The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism.' *Journal of Ethics* 2: 123-42.
- ----. 1997. 'Alternative Possibilities and the Failure of the Counterexample Strategy.' *Journal of Social Philosophy* 28: 71-85.
- Mele, Alfred. 2000. 'Reactive Attitudes, Reactivity, and Omissions.' *Philosophy and Phenomenological Research* 61: 447-452.
- ----. 1995. *Autonomous Agents*. New York: Oxford University Press.
- Mele, Alfred, and David Robb. 1998. 'Rescuing Frankfurt-Style Cases.' *Philosophical Review* 107: 97-112.
- Naylor, Marjory. 1984. 'Frankfurt on the Principle of Alternate Possibilities.' *Philosophical Studies* 46: 249-58.
- Neely, Wright. 1974. 'Freedom and Desire,' *Philosophical Review* 83: 32-54.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA.: Harvard University Press.
- O'Connor, Timothy. 2002. "Free Will", *The Stanford Encyclopedia of Philosophy* (Spring 2002 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2002/entries/freewill/>.
- ----. 2000. *Persons and Causes*. New York: Oxford University Press.
- ----. 1993. 'On the Transfer of Necessity.' *Noûs* 27: 204-218.
- Otsuka, Michael. 1998. 'Incompatibilism and the Avoidability of Blame.' *Ethics* 108: 685-701.
- Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- ----. 1995. 'Determinism al Dente.' *Noûs* 29: 21-45.
- Rowe, William. 1991. *Thomas Reid on Freedom and Morality*. Ithaca: Cornell University Press.
- Russell, Paul. 2002a. 'Critical Notice of John Martin Fischer and Mark Ravizza *Responsibility and Control: A Theory of Moral Responsibility*.' *Canadian Journal of Philosophy* 32: 587-606.
- ----. 2002b. 'Pessimists, Pollyannas, and the New Compatibilism.' In Kane, ed., 2002.
- ----. 1995. *Freedom and Moral Sentiment*. New York: Oxford University Press.
- ----. 1992. 'Strawson's Way of Naturalizing Responsibility.' *Ethics* 102: 287-302.
- Saunders, John Turk. 1968. 'The Temptation of Powerlessness,' *American Philosophical Quarterly* 5: 100-8
- Schlick, Moritz. 1939. 'When is a Man Responsible?' in *Problems of Ethics* translated by David Rynin 143-6. Reprinted in Berofsky, 1966: 54-62.
- Schoeman, Ferdinand, ed. 1987. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge: Cambridge University Press.
- Slote, Michael. 1995. 'Review of Peter van Inwagen's *An Essay on Free Will*.' *Journal of Philosophy* 82.
- ----. 1982. 'Selective Necessity and the Free-Will Problem.' *Journal of Philosophy* 79: 5-24.
- Strawson, Galen. 1986. *Freedom and Belief*. Oxford: Clarendon Press.

- Strawson, P. F. 1962. 'Freedom and Resentment.' *Proceedings of the British Academy* 48: 187-211.
- Stump, Eleonore. 1996a. "Libertarian Freedom and the Principle of Alternative Possibilities." In Howard-Snyder and Jordan, 1996: 73-88.
- ----- 1996b. 'Persons: Identification and Freedom.' *Philosophical Topics*. 24: 183-214.
- ----- 1990. 'Intellect, Will, and the Principle of Alternate Possibilities.' In Beaty, 1990: 254-85.
- Taylor, Richard. 1974. *Metaphysics*. Englewood Cliffs: Prentice Hall.
- van Inwagen, Peter. 2002. 'Free Will Remains a Mystery?' In *The Oxford Handbook of Free Will*. ed. Robert Kane. New York: Oxford University Press, pp. 158-77.
- ----- 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- -----, ed. 1980. *Time and Cause*. Dordrecht: D. Reidel.
- ----- 1978. 'Ability and Responsibility.' *Philosophical Review* 87: 201-24.
- ----- 1975. 'The Incompatibility of Free Will and Determinism.' *Philosophical Studies*. 27:185-99.
- van Straaten, Zak, ed. 1980. *Philosophical Subjects: Essays Presented to P.F. Strawson*. Oxford: Clarendon.
- Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA.: Harvard University Press.
- Watson, Gary. 2001. 'Reason and Responsibility.' *Ethics* 111: 374-94.
- ----- 1987. 'Responsibility and the Limits of Evil: Variations on a Strawsonian Theme.' In Schoeman, 1987: 256-86. Reprinted in Fischer and Ravizza, eds., 1993.
- ----- ed. 1982. *Free Will*. New York: Oxford University Press.
- ----- 1975. 'Free Agency.' *Journal of Philosophy* 72: 205-20. Reprinted in Watson, ed. 1982.
- Widerker, David. 1987. 'On an Argument for Incompatibilism.' *Analysis* 47: 37-41.
- ----- 1995. 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities.' *Philosophical Review* 104: 247-61.
- Widerker, David, and Michael McKenna, eds. 2003. *Moral Responsibility and Alternative Possibilities*. Aldershot, UK: Ashgate Press.
- Wiggins, David. 1973. 'Towards a Reasonable Libertarianism.' In Honderich 1973: 31-62.
- Wolf, Susan. 1990. *Freedom within Reason*. Oxford: Oxford University Press.
- ----- 1987. 'Sanity and the Metaphysics of Responsibility.' In Schoeman, ed., 1987.
- ----- 1981. 'The Importance of Free Will.' *Mind* 90: 386-405. Reprinted in Fischer and Ravizza, eds. 1993.
- ----- 1980. 'Asymmetrical Freedom.' *Journal of Philosophy* 77: 157-66.
- Wyma, Keith. 1997. 'Moral Responsibility and Leeway for Action.' *American Philosophical Quarterly* 34: 57-70.
- Zimmerman, David. 2002. 'Reasons-Responsiveness and the Ownership of Agency: Fischer and Ravizza's Historicist Theory of Responsibility.' 6: 199-234.
- Zimmerman, Michael J. 1988. *An Essay on Moral Responsibility*. Totowa, NJ: Rowman and Littlefield.

### Acknowledgments

For helpful editorial and philosophical advice, I would like to thank Carl Ginet, Ish Haji, Robert Kane, Sean McKeever, Al Mele, Derk Pereboom, Paul Russell, Edward Zalta, and two subject editors of *The Stanford Encyclopedia of Philosophy*, John Fischer and David Velleman.

Copyright © 2004 by  
Michael McKenna